DOCUMENT RESUME

ED 134 938 CS 003 193

AUTHOR	Jenkins, Joseph R.; Pany, Darlene
TITLE	Curriculum Biases in Reading Achievement Tests. Technical Report No. 16.
INSTITUTION	Bolt, Beranek and Newman, Inc., Cambridge, Mass.; Illinois Univ., Urbana. Center for the Study of Reading.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
PUB DATE	Nov 76
CONTRACT	400-76-0116
NOTE	24p.
EDRS PRICE	MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS	Basic Reading; *Beginning Reading; *Curriculum Evaluation; Grade 1; Grade 2; Primary Education; Program Evaluation; Reading Materials; *Reading Research; *Reading Tests; Standardized Tests; *Test Bias
IDENTIFIERS	*Center for the Study of Reading (Illinois)

ABSTRACT

The extent and direction of curriculum bias in standardized reading achievement tests were examined. Bias was estimated by comparing the relative overlap in the contents of five separate reading achievement tests with the contents of seven commercial reading series at first-grade and second-grade levels. Overlap between each achievement test and each reading series is reported in terms of achievement test grade equivalent scores that would be expected, given mastery of the words which appear both as content in a reading series and as achievement test items. Results indicate clear discrepancies between the grade equivalent scores obtained, both between tests for a single curriculum and on a single test for different reading curricula. The implications of the apparent curriculum bias of achievement tests are discussed as they relate to teacher, child, and curriclum evaluation, to reading placement, and to applied educational research. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED134938

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

CENTER FOR THE STUDY OF READING*

Technical Report No. 16

CURRICULUM BIASES IN READING ACHIEVEMENT TESTS

Joseph R. Jenkins and Darlene Pany

University of Illinois at Urbana-Champaign

November 1976

University of Illinois
at Urbana-Champaign
1005 West Nevada Street
Urbana, Illinois 61801

Bolt Beranek and Newman Inc.
50 Moulton Street
Cambridge, Massachusetts 02138

The research reported herein was supported in part by the National Institute of Education under Contract No. MS-NIE-C-400-76-0116.

*Designation as a Center pending approval.

CS 003 193

Abstract

The extent and direction of curriculum bias in standardized reading achievement tests are examined. Bias was estimated by comparing the relative overlap in the contents of five separate reading achievement tests with the content of seven commercial reading series at first and second grade levels. Overlap between each achievement test and each reading series is reported in terms of achievement test grade equivalent scores that would be expected given mastery of the words which appear both as content in a reading series and as achievement test items. Results indicate clear discrepancies between the grade equivalent scores obtained both between tests for a single curriculum and on a single test for different reading curricula. The implications of the apparent curriculum bias of achievement tests are discussed as they relate to teacher, child, and curriculum evaluation, to reading placement, and to applied educational research.

CURRICULUM BIASES IN READING ACHIEVEMENT TESTS [1]

Joseph R. Jenkins

and

Darlene Pany

University of Illinois at Urbana-Champaign

Information furnished by standardized, norm referenced achievement tests influences a broad range of educational decisions. Government agencies use achievement test results to assess the impact of federally supported educational programs. School board budget allocations are modified by achievement test results. Administrators, and sometimes parents, evaluate teachers on the basis of student performance on these tests. School psychologists rely on achievement test results to justify recommendations for special education supportive services. Educators at a number of levels use them to evaluate curricula. Researchers use them to assess the effects of a variety of educational arrangements such as open classrooms, token economies, teacher-pupil ratios, school desegregation, and curricular innovations. And, of course, teachers use standardized, norm-referenced instruments to diagnose children's learning needs, to make placements in a curriculum, and to evaluate student academic growth.

Test developers have openly encouraged consumer confidence in their instruments. Their product endorsements specifically detail the variety of appropriate uses of achievement test results.

Achievement tests:

> "... [tell] what pupils have learned in school" (Metropolitan Achievement Test Primary I Teacher's Handbook, 1970, p. 2),
>
> "... provide a basis for reporting pupils' achievement to parents" as well as permit one "to compare present and past achievement in order to determine and evaluate the rate of progress" (Stanford Achievement Primary I Battery Directions for Administering, 1964, p. 30),
>
> "... [serve as] warning signals [to] give the pupil special help within the framework of the regular instructional program or...request help from various specialists in the school" (Metropolitan Achievement Test Primary I Teacher's Handbook, 1970, p. 13),
>
> "... [permit] the determination of instructional levels in school children" and "the assignment of children to instructional groups..." (Wide Range Achievement Test Manual, 1965, p. 1),
>
> "... are a source of information on which to base major curriculum changes" (Stanford Achievement Test Primary I Battery Directions for Administering, 1964, p. 30).

In spite of the fact that achievement tests are highly touted by their developers and publishers, they are not without their critics. Objections to conventional achievement tests have generally taken one of two forms. Advocates of criterion-referenced testing argue that norm-referenced measures tell little about what an individual child has learned or not learned (Carver, 1972). Instead, norm-referenced tests can indicate only how a particular child's score compares with scores obtained by children in the norming sample (Popham, 1974). Others have criticized achievement tests on the basis of research which indicates that achievement test performance often fails to correspond with performance in actual classroom curricula (Brown, Note 1; Glaser, Note 2; McCracken, 1962; Sipay, 1964). Carver has concluded that "... grade level scores on reading tests have no connection with grade level difficulty of basal

readers or other curriculum materials" (Carver, 1972, p. 300). Eaton and Lovitt (1972) have furthermore presented data which raises doubts about the capacity of achievement tests to measure children's annual academic growth.

Despite growing suspicion of conventional achievement tests in some circles, the educational community continues to place enormous confidence in them; when achievement test results run counter to teachers' perceptions of children's progress, the achievement test score is usually accepted as the more valid assessment. When a child receives a low score on a test, it is the child, the teacher, and/or the curriculum that is blamed. Unprepared children, inadequate curricula, and unsystematic teachers are definitely plausible explanations for poor test performance. However, there is another explanation that is rarely considered, namely that achievement tests may not measure what was taught. The present investigation focuses attention on this latter interpretation and examines the extent to which reading achievement tests may not adequately sample particular instructional programs, even though the instructional programs may themselves be adequate.

Most conventional reading achievement measures are composed of one or more subtests such as word recognition, vocabulary meaning, and comprehension. Each of these tests are, in turn, composed of a particular set of words that the child must be able to read. Test developers assure the consumer that the test items (words) are a representative sample of words taught in a wide variety of reading curricula (Sort Directions, 1963; PIAT Manual, 1970; Metropolitan Achievement Test Primary I Teacher's Handbook, 1970). Despite these assurances, it is entirely possible that

the sample of words appearing on a reading achievement test overlaps the words taught in one curriculum more than those taught in another. Reading tests could, in fact, be positively or negatively biased toward a specific reading curriculum by virtue of the particular sample of test words. Such biases might be detected by determining the overlap between various reading curricula and various achievement measures.

The authors recognize that content overlap between reading curricula and achievement tests is not the only factor which determines how children taught in a particular curriculum will perform on an achievement test. In some instances, children will correctly identify on a test words which were not directly taught in their reading curriculum; they may have learned words from sources other than their reading program (e.g., television, family members and peers). They also may decode some unfamiliar test words by applying phonic rules that were taught directly (synthetic phonics) or indirectly (analytic phonics) (Chall, 1967). On the other hand, children may fail to read some words on an achievement test, even though the words were included in their reading program. They may not have mastered those words in the first place, or at the time of the test they may have forgotten words that they once knew. In spite of the fact that performance may reflect factors other than the reading curriculum, it seems safe to assume that the content words of a reading curriculum make the single, largest contribution to a child's reading vocabulary. Factors other than the content words themselves might be expected to counterbalance one another so that reported grade equivalent test scores, estimated solely from curriculum content words, should reasonably indicate both the extent and direction of curriculum bias in selected reading achievement tests.

Just how fair are normalized test scores when related to specific reading curricula? How much weight should be placed on those test results in terms of evaluating or placing students, and communicating information about individual or group achievement in a given school year? To address these questions, the authors assessed the extent and direction of curriculum bias in five widely employed standardized achievement tests: the Wide Range Achievement Test (WRAT); the Peabody Individual Achievement Test (PIAT); the Metropolitan Achievement Test (MAT); the Stanford Achievement Test (SAT); and the Slosson Oral Reading Test (SORT). Bias was estimated by comparing the relative overlap in the contents of these different reading achievement tests with the first and second grade contents of seven commercial reading series: Economy (Keys to Reading); Ginn (Reading 360); Macmillan (The Bank Street Readers); Macmillan (Macmillan Reading Program); Houghton-Mifflin (Reading for Meaning); Science Research Associates (The SRA Reading Program); and, McGraw-Hill (Sullivan Associates Programmed Reading).

Method

First and second grade books from seven basal reading series were surveyed (see Table 1). Publisher's guidelines were used to determine which books in a series corresponded to first and second grade content. Teachers'

- -

Insert Table 1 about here

- -

manuals were used to compile alphabetized word lists for each book in a series. Unless specifically indicated as "supplementary" (Houghton-Mifflin), 'enrichment" (Ginn), or "sounding vocabulary" (Economy), all words were assumed to appear in the reading text and to be taught for mastery.

Next, alphabetized lists of all words in seven standardized tests and subtests of word recognition were prepared. In all but two instances, reading tests and subtests which involved sentence or paragraph reading were excluded; the exceptions were the MAT Primary II Word Knowledge Subtest and the SAT Primary I and Primary II Paragraph Meaning Subtests. For these tests, a list was made only of those words which were correct responses.

The extent of overlap between each reading series and each achievement test could then be assessed by comparing test word lists with curriculum word lists to determine the total number of word matches per grade level. For example, of the 50 words taught in Economy, Level 2, (the first of five books read in first grade) three words, "jump," "play," and "run" appear on the PIAT Word Recognition Subtest. Thus, Economy, Level 2, and the PIAT yield three word matches. Only words which appeared in the same form both on the test and in the curriculum were counted as matches. Exceptions included words with -s, -d, -ed, and -ing endings, which did not change the root word. The words "walk" and "walking" would qualify as matches, but the words "ride" and "riding" would not, since the "e" is dropped in "riding." Similarly, the words "hunger" and "hungry" would not qualify as matches.

The PIAT, WRAT and SORT all have error ceilings, which if reached, conclude testing (see Table 2). Thus, it was necessary to consider the sequence of test words, when locating word matches. Some potential word matches were excluded since the error ceiling would have terminated testing before the word appeared.

- -

Insert Table 2 about here

- -

Since word recognition tests are scored by one point for each correct word, the total number of word matches yielded a raw score. Raw scores were then converted to grade equivalent scores according to test manual specifications. For example, comparison of words from the first grade level of the SRA Basic Reading Series (Book A-D) with the words appearing on the SORT indicated 20 word matches (raw score = 20), yielding a grade equivalent score of 1.0, according to the following calculation: 3 (Book A words appearing on the SORT) + 2 (Book B words) + 8 (Book C words) + 7 (Book D words) = 20 (Raw score), converted to grade equivalent = 1.0.

Two people independently matched reading test and curriculum word lists. A third person compared lists of matching words and reconciled disagreements. Raw, standardized and grade equivalent scores were also computed by two persons independently.

Results and Discussion

Table 3 summarizes the overlap between each achievement test and each reading series, for first and second grade levels. The overlap is reported in terms of achievement test grade equivalent scores that would be expected, given the words which appear both as items on an achievement test and as instructional content in a reading series.

- -

Insert Table 3 about here.

- -

Inspection of Table 3 reveals clear discrepancies between the grade equivalents obtained both between tests for a single curriculum and on a single test for different reading curricula. The extent of curriculum

bias is not uniform across all achievement tests. At the first grade level, the MAT appears to exhibit the least curriculum bias in that scores from all seven reading curricula fall within a narrow range (0.4 grade equivalences), compared to a range of 1.2 grade equivalent scores obtained on the SORT. However, the MAT's consistently low grade equivalent scores raises doubts as to the test's accuracy in describing actual grade level achievement. Certain of the reading curricula seem to enjoy consistently high overlap with all five achievement tests. At the first grade level, the Economy Series obtains or ties with another curriculum for the highest grade equivalent on five out of the seven tests and subtests examined in this study. Likewise, some curricula at the first grade level obtain low grade equivalents across several tests. Ginn 360 obtains or ties for the lowest first grade equivalent score on six of seven tests and subtests, Houghton-Mifflin on two of seven, and SRA on two of seven. Some implications of the apparent biases between achievement tests and reading curricula are explored be below.

Student Evaluation

Students, teachers, and curricula are all subject to evaluation based on standardized test scores. For a particular student, the scores are often used to measure the amount of growth over some period. Children making "normal" progress are expected to advance one full grade equivalent for each year spent in school. Examination of the scores (Table 3) for any curriculum, however, reveals that the amount of growth varies depending upon the particular test employed. Hypothetically, a child who learned the content words in Grade 1 of Houghton-Mifflin by the end of Grade 2 would gain one year and four months according to the PIAT, one year and two months in Word Knowledge (MAT), zero in Word Analysis (MAT), eight

months according to the SORT, seven months in Paragraph Meaning (SAT), and only four months on the WRAT.

An equally distorted picture is presented for Sullivan curriculum growth as measured by the five tests. Using the PIAT scores, a child would be judged as "average" at the end of first grade, but by the end of second grade would have gained only four months. Although 16 months growth is shown for the third grade, it still appears that the child would enter fourth grade behind grade level. If growth in that same curriculum is measured with the SORT, only one month's growth is indicated by the end of Grade 1. That can be contrasted with second and third grades when 14 and 18 months progress is possible. If the SORT were substituted for the PIAT, the child completing three years of the Sullivan curriculum will begin fourth grade somewhat above grade level. Which test results should be believed in evaluating the child's progress? It appears that measured progress may be more reflective of test and curriculum combinations than of teaching and learning. A second grade teacher using SRA might "produce" a child reading at or above grade level by the end of second grade merely by selecting the WRAT or PIAT instead of the MAT. If dramatic "growth" is desired, s/he could use the SORT and obtain 19 months gain from the end of first to the end of second grade.

Provision of special education and other supportive services are based to a significant degree on children's achievement test performance. For a child to be classified as educable mentally handicapped or learning disabled in most states, s/he must score below his/her grade expectancy on an achievement test. Low achievement test performance is also used as corroborative

evidence for emotional disturbance. Federal Title I and Title VII guidelines include an achievement criterion in identifying candidates for services. How many times have recommendations for retention or special class placement been prompted or supported by distorted test results? Often, decisions made during a staffing about a child's educational placement are based on normative data, with achievement test results serving as the primary source of information.

Curriculum Evaluation

In addition to using achievement test results to measure pupil growth, administrators might use achievement test results as a means of evaluating a particular curriculum (Stanford Achievement Test Primary I Battery for Administering, 1964; Metropolitan Achievement Test Primary I Teacher's Handbook, 1971.) Suppose that a school district accepted this suggestion, and field tested several reading series in different classrooms for two years before deciding which reading series to adopt. If the school district used selected students' scores on the PIAT to evaluate the different reading series, they would probably select the Economy or Houghton-Mifflin texts. If, instead, they used the WRAT to evaluate growth, they would probably choose SRA, Economy or Bank Street Readers. It is highly doubtful that conventional achievement tests can serve as unbiased estimates of a curriculum's worth, at least at the early grade levels. Perhaps, at a time when all word recognition skills should have been mastered (e.g., Grade 4), then an achievement test would not be seriously biased toward any particular curriculum, at least by virtue of the vocabulary it contains. However, other sources of bias exist. These will be discussed later.

Teacher Evaluation

Like children and curricula, teachers may also be subject to evaluations which employ standardized test results. With the emphasis on accountability in education, a teacher's ability may be judged by the number of children in his/her class who are at or near grade level according to year-end achievement tests. A first grade teacher using the Sullivan Programmed Readers would be rated highly by PIAT or WRAT results. That same teacher might appear quite inadequate if the MAT, SORT, or SAT were used. Similarly, what can be said in defense of the second grade teacher using Sullivan whose class "gained" only four months on the PIAT. The widespread reluctance of teachers to be held accountable for their performance may be justified, especially if their effects are measured by biased instruments. The right combination of curriculum and achievement test may enhance the "effects" of a poor teacher, whereas the unfortunate combination of test and curriculum may penalize a good teacher.

Reading Placement (Diagnosis)

Achievement test grade equivalent scores are useful, according to some authorities, in placing children within a reading series. According to the results reported in Table 3 one sees that the accuracy of placement decisions is greatly affected by the combination of achievement test and reading curriculum. A student who mastered first grade Ginn vocabulary would obtain a SORT grade equivalent of 1.4. The same student, if he had read in Economy, could expect a score of 2.2. What is the proper interpretation of these results? Do they indicate that a student using Ginn is not really prepared to read second grade material, but that one using Economy is?

The inadequacy of achievement test results for making placement decisions can be further illustrated. Suppose that a child is new to a school as a second grader. In September his new teacher administers the SORT so that a placement in Macmillan can be made. The child, having read Books 1-7 of the Sullivan Programmed Readers at his former school scores a grade equivalent of 1.1. Other students in the class (who finished the Macmillan first grade readers) receive on the average a grade equivalent of 1.9, close to grade level. The teacher might conclude that the new child is a non-reader, and that s/he will not "fit" with the rest of his/her second grade. The teacher might request supportive services for the new student, or possibly consider a special education placement. However, if the same child were given a WRAT, a grade equivalent of 2.0 would indicate that he, too, is reading at grade level, and is only a little behind his classmates who, given their Macmillan background, could be expected to obtain a WRAT score of 2.3. In this case the teacher would probably assume the child could safely be placed in a "middle" reading group, beginning a 2-1 reader.

A teacher relying on grade equivalent scores to make a placement decision in a particular curriculum may be led to radically different conclusions depending upon the selection of achievement test and the child's previous reading curriculum. All a teacher knows after administering a standardized test is how many words on that particular test a child knows, and how that score compares to other children in the class, and to some children on whom the test was normed. What the teacher does not know is which words a child can read in a particular reading series. It is that information which is needed to place a child at an appropriate instructional level in a given curriculum.

Educational Research

Applied researchers in education have understood that in order to assess the relative effects of any independent variable on student achievement, all other variables which could conceivably influence student achievement must be controlled. Many studies, particularly those conducted in normal school settings, have appeared in which the independent variable under study (e.g., teacher-pupil ratios; classroom organization; type of special education services, etc.) is confounded with different classroom curricula. In some research reports, the authors do not feel compelled even to mention whether curricula were controlled across treatments. The assumption that achievement tests were unbiased samples of commercial curricula is, apparently, responsible for the failure to control carefully the curriculum used by different treatment groups. The results of the present investigation would suggest that conclusions drawn from any study where the dependent variable was student achievement measured by conventional instruments would be significantly attentuated, unless the classroom curriculum was carefully controlled across treatment conditions. Inconsistent findings from study to study, so familiar in the education literature, may in part be accounted for by uncontrolled but systematic biases between curricula and achievement tests.

Conclusions

The data from the present investigation strongly suggest that a basic assumption underlying standardized achievement measures, that they representatively sample different curricula, cannot reasonably be held; clear, significant biases exist. The nature of this bias is such that student achievement in a particular curriculum may in no way be reflected

by achievement test scores. Such biases must be acknowledged and considered any time that a standardized, norm-referenced achievement test is used for decision making.

In all likelihood, achievement test bias extends beyond measurement of single word recognition skill. Reading comprehension tests also require children to read a specific set of words and respond to them in some fashion (e.g., recall particular facts, draw an inference or supply missing words). If words that compose test items on a reading comprehension tests are more congruent with one reading curriculum than with another, then children's performance on such test(s) may also be affected. To compound matters, reading comprehension tests include additional sources of potential bias such as question format (e.g., cloze vs. multiple choice) sentence construction (e.g., multiple vs. complex) and topic (e.g., baseball vs. sewing) over and above those sources of bias found in word recognition measures. Thus, the problem of achievement test bias does not conveniently disappear when reading comprehension tests are substituted for word recognition tests; instead, the problem grows.

What educators need is an instrument to measure learning that is sensitive to curricular differences. Some form of criterion-referenced or curriculum-based assessment may provide the solution. Frequent and direct measures of a child's performance in a specific curriculum should reveal what skills within the curriculum have or have not been mastered, as well as provide some index of progress which would be sensitive to what was being taught.

Footnote

1. The authors are grateful to Laura Aull and Kurt Pany for their invaluable technical assistance and to Barbara Wilcox and Judith Arter for their comments on an earlier version of the manuscript.

Reference Notes

1. Brown, S. A comparison of five widely used standardized reading tests and an informal reading inventory for a selected group of elementary school children. Unpublished doctoral dissertation, University of Georgia, 1963.

2. Glaser, N. A comparison of specific reading skills of advanced and retarded readers of fifth grade reading achievement. Unpublished doctoral dissertation, University of Oregon, 1964.

Test References

Metropolitan Achievement Test, Primary I and II Battery. W. Durost, H. Bixler, J. W. Wrightstone, G. Prescott & I. Balow. New York: Harcourt, Brace, Jovanovich, 1970.

Peabody Individual Achievement Test, Volume I: Reading Recognition. L. Dunn & F. Markwardt. Circle Pines, MN: American Guidance Service, 1970.

Slosson Oral Reading Test. R. Slosson. East Aurora, NY: Slosson Educational Publications, 1963.

Stanford Achievement Test, Primary I and II Battery. T. Kelly, R. Madden, E. Gardner & H. Rudman. New York: Harcourt, Brace, and World, 1964.

Wide Range Achievement Test. J. Jastak, S. Bijou & S. Jastak. Wilmington, DE: Guidance Associates, 1965.

References

Black, I. S. The Bank Street Reading Series. New York: Macmillan, 1965.

Buchanan, C. D. Sullivan Associates Programmed Reading (Series 2, revised edition). New York: McGraw-Hill, 1968.

Carver, R. Reading tests in 1970 vs. 1980: Psychometric vs. edumetric. Reading Teacher, 1972, 26, 299-302.

Chall, J. Learning to read: The great debate. New York: McGraw-Hill, 1967.

Clymer, T. Reading 360. Boston: Ginn and Co., 1969.

Dunn, L., & Markwardt, F. Peabody Individual Achievement Test Manual. Circle Pines, MN: American Guidance Service, 1970.

Durost, W., Bixler, H., Wrightstone, J. W., Prescott, G., & Balow, I. Metropolitan Achievement Test Primary I Teacher's Handbook. New York: Harcourt, Brace, Jovanovich, 1970.

Eaton, M., & Lovitt, R. Achievement tests vs. direct and daily measurement. In G. Semb (Ed.), Behavior analysis and education. University of Kansas, 1972.

Farr, R. Reading: What can be measured? Newark, DE: International Reading Association, 1969.

Harris, A., & Clark, M. The Macmillan Reading Program. New York: Macmillan, 1970.

Harris, T., & Creekmore, M. Keys to reading. Oklahoma City: Economy, 1972.

Jastak, J. F., & Jastak, S. R. The Wide Range Achievement Test Manual. Wilmington, DE: Guidance Associates, 1965.

Kelly, T., Madden, R., Gardner, E., & Rudman, H. Stanford Achievement Test Primary I Battery Directions for Administering. New York: Harcourt, Brace, and World, 1964.

McCracken, R. A. Standardized reading tests and informal reading inventories. Education, 1962, 82, 366-69.

McKee, P., Harrison, M. L., McCowen, A., Lehr, E., & Dunn, W. Reading for meaning (fourth edition). New York: Houghton-Mifflin, 1966.

Popham, W. J. An approaching peril: Cloud-referenced tests. Phi Delta Kappan, 1974, 55, 614-615.

Rasmussen, D., & Goldbert, L. The SRA Reading Program: Basic Reading Series. Chicago: Science Research Associates, 1970.

Sipay, E. A comparison of standardized reading scores and functional reading levels. The Reading Teacher, 1964, 17, 265-68.

Table 1

Reading Curricula Grade Levels

Series (Publisher)	Level	Grade
The Bank Street Readers (Macmillan 1965)	Preprimers (2)	1
	Primer (Around the City)	1
	1-1 (Uptown, Downtown)	1
	2-1 (My City)	2
	2-2 (Green Light Go)	2
Keys to Reading (Economy 1972)	2 (Pug)	1
	3 (Sun Tree)	1
	4 (Zip! Pop! Go!)	1
	5 (Green Feet)	1
	6 (Blue Dilly Dilly)	1
	7 (Curbstone Dragons)	2
	8 (Mustard Seed Magic)	2
Reading 360 (Ginn and Co., 1969)	2 (My Sound and Word Book)	1
	3 (A Duck is a Duck)	1
	4 (Helicopters and Gingerbread Men)	1
	5 (May I Come In?)	1
	6 (Seven is Magic)	2
	7 (The Dog Next Door and Other Stories)	2
Reading for Meaning (Houghton-Mifflin 1966)	Preprimers (3)	1
	Primer (Jack and Janet)	1
	1-1 (Up and Away)	1
	2-1 (Come Along)	2
	2-2 (On We Go)	2
Macmillan Reading Program - Primary Grades (Macmillan 1970)	Preprimers (3)	1
	Primer (Worlds of Wonder)	1
	1-1 (Lands of Pleasure)	1
	2-1 (Enchanted Gates)	2
	2-2 (Shining Bridges)	2
The SRA Reading Program (Science Research Associates 1971)	A (A Pig Can Jig)	1
	B (A Hen in a Fox's Den)	1
	C (Six Ducks in a Pond)	1
	D (A King on a Swing)	1
	E (Kittens and Children)	2
	F (The Purple Turtle)	2
	G (Tony's Adventure)	2
Sullivan Associates Programmed Reading (McGraw-Hill 1968)	Primer	1
	Books 1 through Book 7	1
	Books 8 through Book 14	2
	Books 15 through Book 21	3

Table 2

Scoring Criteria

Wide Range Achievement Test (WRAT)

Error ceiling:	12 consecutive errors in word recognition
Raw score:	number of correct words plus 25
Assumptions:	child can identify 13 letters of the alphabet, match ten identical letters, and can identify two letters in his/her name (25 points)

Peabody Individual Achievement Test (PIAT)

Error ceiling:	five errors in seven consecutive words
Raw score:	error ceiling word number minus total number of errors
Assumptions:	a child can identify letter names and can match identical letters, words, and pictures (18 points); starting point (basal level) is the first word (item number 19)

Metropolitan Achievement Test, Primary I and Primary II (Form F) (MAT) and Stanford Achievement Test, Primary I and Primary II (Form W) (SAT)

Error ceiling:	none
Raw score:	number correct
Standardized score:	conversion table provided in test manual
Grade equivalent score:	conversion table from standardized score provided in test manual

Primary I was used to calculate grade one scores.
Primary II was used to calculate grade two scores. First as well as second grade words were matched to the Primary II test words.
The MAT Primary II Word Knowledge subtest and SAT Primary I and Primary II Paragraph Meaning subtests are multiple choice tests which involve reading a sentence or paragraph. Since a curriculum actually may not include the words that must be read to select the correct word answer, scores on those tests appearing in Table 3 may be inflated.

Slosson Oral Reading Test (SORT)

Error ceiling:	100% incorrect words in a column of 20 words
Raw score:	number correct
Grade equivalent:	Raw score divided by two (table provided in directions)

Table 3

Grade Equivalent Scores Determined by Matching Specific Reading Text Words to Standardized Reading Test Words

Curriculum	PIAT	Test MAT		SORT	SAT		WRAT
		Word Knowledge	Word Analysis		Word Reading	Paragraph Meaning	
Bank Street Readers							
Grade 1	1.5	1.0	1.1	1.8	1.0	1.6	2.0
Grade 2	2.8	2.5	1.2	2.9	-	2.9	2.7
Economy*							
Grade 1	2.0 (3.3)	1.4 (1.9)	1.2 (1.8)	2.2 (2.7)	1.5 (1.7)	1.8 (2.3)	2.2 (3.0)
Grade 2	3.3 (3.8)	1.9 (3.7)	< 1.0 (2.0)	3.0 (3.5)	- -	2.6 (3.2)	3.0 (3.6)
Ginn 360**							
Grade 1	1.5 (1.5)	< 1.0 (< 1.0)	< 1.0 (1.0)	1.4 (1.5)	< 1.0 (< 1.0)	1.5 (1.5)	1.7 (1.8)
Grade 2	2.2 (2.8)	2.1 (2.5)	< 1.0 (1.1)	2.7 (2.7)	- -	1.9 (1.9)	2.3 (2.5)
Houghton-Mifflin***							
Grade 1	2.0 (3.1)	1.1 (1.9)	< 1.0 (1.7)	1.6 (2.9)	< 1.0 (1.7)	1.6 (2.4)	2.0 (3.1)
Grade 2	3.4 (3.8)	2.3 (2.9)	< 1.0 (2.0)	2.4 (3.4)	- -	2.3 (3.4)	2.4 (2.9)
Macmillan							
Grade 1	1.8	1.1	1.0	1.9	1.1	1.6	2.3
Grade 2	2.2	2.5	1.1	2.9	-	2.5	2.6
SRA							
Grade 1	1.5	1.2	1.3	1.0	1.0	1.6	2.1
Grade 2	3.1	2.5	1.4	2.9	-	2.9	3.5
Sullivan							
Grade 1	1.8	1.4	1.2	1.1	1.2	1.6	2.0
Grade 2	2.2	2.4	1.1	2.5	-	3.1	2.5
Grade 3	3.8	-	-	4.3	-	-	3.5

*Scores in parentheses reflect the inclusion of words listed as "sounding vocabulary."
** " " " " " " " " " " "enrichment."
*** " " " " " " " " " " "supplementary."